

The Dynamic Double Flash Illusion: Auditory Triggered Replay of Illusory Visual Expansion

Noelle R. B. Stiles^{1,2,*}, Armand R. Tanguay, Jr.^{2,3} and Shinsuke Shimojo²

¹ Department of Ophthalmology, University of Southern California, Los Angeles, CA, USA

² Division of Biology and Biological Engineering, and Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA, USA

³ Departments of Electrical Engineering-Electrophysics, Chemical Engineering and Materials Science, Biomedical Engineering, Ophthalmology, and Physics and Astronomy; Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, USA

Received 21 December 2018; accepted 17 July 2019

Abstract

In the original double flash illusion, a visual flash (e.g., a sharp-edged disk, or uniformly filled circle) presented with two short auditory tones (beeps) is often followed by an illusory flash. The illusory flash has been previously shown to be triggered by the second auditory beep. The current study extends the double flash illusion by showing that this paradigm can not only create the illusory repeat of an on-off flash, but also trigger an illusory expansion (and in some cases a subsequent contraction) that is induced by the flash of a circular brightness gradient (gradient disk) to replay as well. The perception of the dynamic double flash illusion further supports the interpretation of the illusory flash (in the double flash illusion) as similar in its spatial and temporal properties to the perception of the *real* visual flash, likely by replicating the neural processes underlying the illusory expansion of the real flash. We show further that if a gradient disk (generating an illusory expansion) and a sharp-edged disk are presented simultaneously side by side with two sequential beeps, often only one visual stimulus or the other will be perceived to double flash. This indicates selectivity in auditory–visual binding, suggesting the usefulness of this paradigm as a psychophysical tool for investigating crossmodal binding phenomena.

Keywords

Crossmodal, multisensory, vision, audition, illusion, double flash, auditory–visual binding

* To whom correspondence should be addressed. E-mail: nstiles@caltech.edu

1. Introduction

Audition has been found to alter visual perception, particularly over short time scales and in the visual periphery (Shimojo and Shams, 2001). The double flash illusion enlists both of these sensory domains to induce a single visual flash to be perceived twice when paired with two brief auditory tones (beeps) (Shams *et al.*, 2000, 2002; Shimojo and Shams, 2001) (Fig. 1). In particular, if a peripherally-located sharp-edged disk (uniformly filled circle) is flashed (nearly) simultaneously with a short beep, then a following second beep is often perceived to be accompanied by an illusory flash. Electroencephalography (EEG) measurements have been used with the double flash paradigm to determine that activation in early visual cortex is correlated with (and likely generates the perception of) the illusory flash (Mishra *et al.*, 2007; Shams *et al.*, 2001, 2005; Watkins *et al.*, 2006). In addition, MEG (magnetoencephalography) measures have shown that both crossmodal connectivity and phase

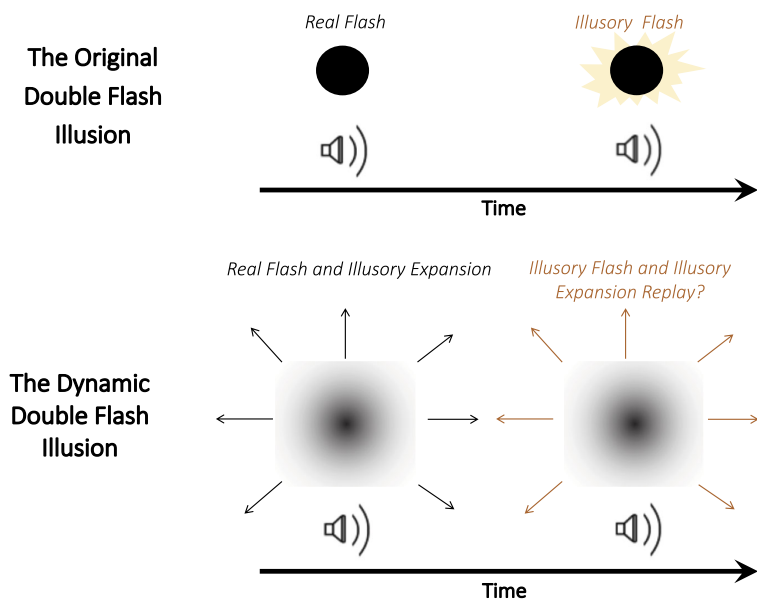


Figure 1. The Double Flash and Dynamic Double Flash Illusions. A schematic diagram of the original double flash illusion experimental design is shown (top) with a sharp-edged disk flash accompanied by two beeps. These beeps trigger a second flash, which is illusory (top, far right). A schematic diagram of the dynamic double flash illusion is displayed at the bottom of the figure. In this case, a gradient disk is flashed that generates illusory expansion (or expansion and contraction), and two beeps are played as before. An illusory flash (bottom, far right) is perceived following the real flash, in which illusory dynamics are perceived that are similar to the illusory dynamics of the real flash.

synchrony before the illusion predict the probability of the double flash illusion occurring (Keil *et al.*, 2013; Lange *et al.*, 2011). Therefore, the illusory flash is currently understood to be generated by activity in auditory cortex that induces the reactivation of early regions of the visual processing pipeline.

To probe the similarity of the illusory flash to the real flash, we employed a visual spatial gradient disk that generates illusory expansion (and sometimes also contraction) when flashed, likely due to contrast-dependent motion processing mechanisms (Seiffert and Cavanagh, 1999) (Fig. 1). We show in Experiment 1 of this paper that the illusory dynamics of the gradient disk are replicated in the illusory flash of the double flash paradigm. This indicates that the illusory flash has similar visual processing and visual perceptual properties to those of the real visual flash, further suggestive of its visual-cortical origin.

In addition to clarifying the perception and processing of the illusory flash, the dynamic double flash illusion can be used as a tool to further understand the mechanisms of auditory–visual binding. If a visual stimulus is perceived to double flash, then it can be considered effectively bound to the auditory tones (beeps) that are presented with it. As a consequence, the double flashing of a visual stimulus can be framed as an indicator of auditory–visual binding. Using this metric, in Experiment 2 we compare the auditory–visual binding (i.e., the number of illusory flashes perceived) of two visual stimuli (a sharp-edged disk and a gradient disk) presented simultaneously side by side. This experimental design allows us to test whether auditory–visual binding can be differentially affected by within-vision object grouping. The results of Experiment 2 indicate that one visual stimulus can be selected for stronger auditory–visual binding than its neighbor, even within the short period that the double flash illusion is processed. For purposes of this paper, we define auditory–visual ‘binding’ as an instantaneous perceptual phenomenon that excludes slower cognitive matching processes such as crossmodal correspondences (Spence, 2011).

In this paper, we describe the use of the double flash paradigm to examine the visual processing of illusory flash dynamics (Experiment 1), and to determine the visuospatial properties that optimize auditory–visual binding (Experiment 2). In other words, we will investigate crossmodal integration over short time scales by harnessing the illusory visual dynamics of spatial gradients.

2. Methods

2.1. Participants

Twenty-one participants (12 male and 9 female) took part in the two experiments altogether, with each participant taking part in only one of the two

experiments. Experiment 1 involved ten participants (5 male and 5 female). Two of these participants were included in the number of flashes perceived analysis, but were excluded from the type of perception analysis, due to the fact that they perceived double flashes in less than 5 trials for either the sharp-edged disk or the gradient disk. This partial exclusion enables accurate reporting of the number of flashes perceived by all participants, but obviates the problem of including participants with only a small number of reported double (illusory) flashes in analyzing the type of perception. Experiment 2 involved eleven participants (7 male and 4 female); all of the Experiment 2 participants were different individuals than those involved in Experiment 1.

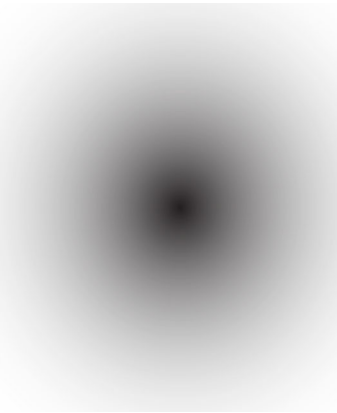
For each experiment, participants were given instructions for only the task they needed to complete (e.g., count the number of flashes), but were not made aware of the experimental goal. Experiments were approved by the Caltech Committee for the Protection of Human Subjects, and all participants gave informed written consent.

2.2. Stimuli

Participants were seated such that their eyes were located approximately 20 inches (50.8 cm) away from an Apple 15 inch Retina MacBook Pro laptop with a native display resolution of 2880×1800 pixels with a pixel density of 220 pixels/inch (set within MATLAB while running the experiment to 1440×900 pixels with a pixel density of 110 pixels/inch) and a pair of built-in speakers. Experiment 1 included a fixation cross (Psychtoolbox-3, screen 'DrawLine' function, [200 200 200] relative brightness) near the top of the screen and a visual stimulus centered 3.5 inches (8.9 cm; 9.9° visual angle) below the fixation cross. The visual stimuli were either a black sharp-edged disk (1 inch in diameter; 2.5 cm; 2.9° ; 110 pixels) or a gradient disk (black center fading to white, originally created at 1531×2606 pixels as shown in Fig. 2 and then scaled to 2.78 inches (7.1 cm; 8.0° ; 306 pixels) high by 4.74 inches (12.0 cm; 13.5° ; 521 pixels) wide) for display. Both visual stimuli were presented on a white background (Psychtoolbox-3, screen 'OpenWindow' function, [255 255 255] relative brightness). The gradient disk was generated in Adobe Illustrator using the gradient tool; the relative brightness profile of the gradient is presented in Fig. 2 (expressed in brightness levels from 0 to 255; corresponding illuminance levels are given below). The gradient disk stimulus vertical profile was truncated at nearly the points of saturation in order to provide adequate separation from the fixation cross.

Experiment 2 also included a fixation cross near the top of the screen and two visual stimuli centered 3.5 inches (8.9 cm; 9.9°) below the fixation cross. The two visual stimuli were 6 inches (15.2 cm; 17.1°) apart (center-to-center), and were either two black sharp-edged disks (1.6 inches in diameter; 4.1 cm; 4.6° ; 176 pixels), two gradient disk stimuli (scaled to 5.56 inches (14.1 cm;

A



B

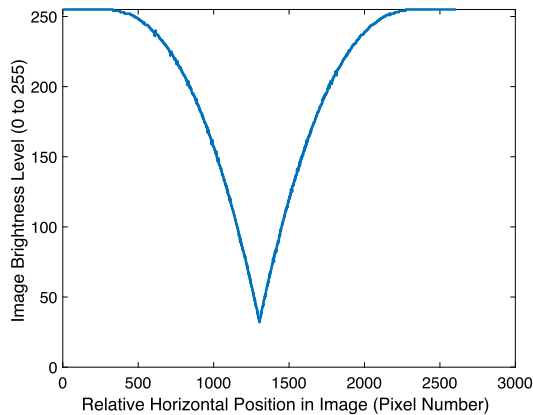


Figure 2. Gradient Disk Stimulus. A circular gradient visual stimulus (gradient disk) was used to generate illusory dynamics in Experiments 1 and 2. The visual stimulus is displayed in Panel A of this figure. The size of the gradient disk image, and its position relative to the fixation cross, are detailed in the Methods section. In Panel B, a horizontal cross-section of the gradient disk is plotted and displayed. The y -axis is the image brightness level (0 to 255; larger numbers indicate higher brightness) and the x -axis is the horizontal position in pixels within the original image itself. The illuminance of the gradient disk as displayed, and its scaling in visual angle for presentation in each of the two experiments, are given in the Methods section.

15.8°; 612 pixels) high by 9.47 inches (24.1 cm; 26.6°, 1042 pixels) wide, partially overlapping at the screen center), or a black sharp-edged disk and a gradient disk (same sizes as above). The gradient disk stimuli were scaled such that they appeared perceptually in pilot studies to be approximately the same ‘size’ as the sharp-edged disks, while maintaining steep enough gradients to produce perceptual dynamics.

The experiment was performed in a small room that was dimly lit. The background illumination was 0.7 lux, as measured with a Gossen MAVOLUX

Model 5032C/B USB Digital Luxmeter. The screen brightness was set to maximum, corresponding to a setting of $[R\ G\ B] = [255\ 255\ 255]$, and was measured to have an illuminance of 245 lux with the luxmeter probe in proximity contact with the screen. The screen background luminance Y was measured with a Konica Minolta CS-100A Chroma Meter to be 134 cd/m^2 , with xy chromaticity coordinates of 0.318, 0.339 (essentially white). The luminance at the center of the sharp-edged disk was measured to be 2.7 cd/m^2 , and the luminance at the center of the gradient disk was measured to be 25.1 cd/m^2 , averaged over a small region centered on the valley of the gradient (approximately 1° of visual angle).

The timing parameters for the visual and auditory stimuli were measured with an oscilloscope after the completion of the experiment. The oscilloscope was configured to display the output from a photodetector circuit that was proximity coupled to the computer display for flash timing in one channel, while simultaneously displaying the audio output from the computer for audio timing in another channel. As such, the stimulus onset, amplitude, and duration for each modality were presented in two synchronized traces (similar to Fig. 3A), triggered by the onset of the first audio beep. The timing for each stimulus type (1F (flash), 1B (beep); 1F, 2B; 1F, 3B) was measured five times, and the time parameters were then averaged across the five measurements. Measured average onset, offset, and duration time parameters for the sharp-edged disk and the gradient disk stimuli are all presented in Table 1. The visual stimuli (sharp-edged disk and/or gradient disk) for Experiments 1 and 2 were presented for approximately 35 milliseconds (see Table 1 for detailed measurements). One, two, or three brief auditory tones (beeps) were paired with the visual stimulus in Experiments 1 and 2; each beep was about 6 milliseconds in duration with a primary frequency of 2.17 kHz and a sampling frequency of 8192 samples per second. The audio (beep) and visual (flash) timing diagrams are detailed in Fig. 3A, and the relative amplitude of the auditory beeps as a function of time is shown in Fig. 3B.

The ambient noise level in the room was approximately 51.7 dB, and the amplitude of each beep was measured to be approximately 59.1 dB, or 7.4 dB above the background noise level. Audio level measurements were taken with an Extech Model 407730 Digital Sound Level Meter.

2.3. Participant Tasks

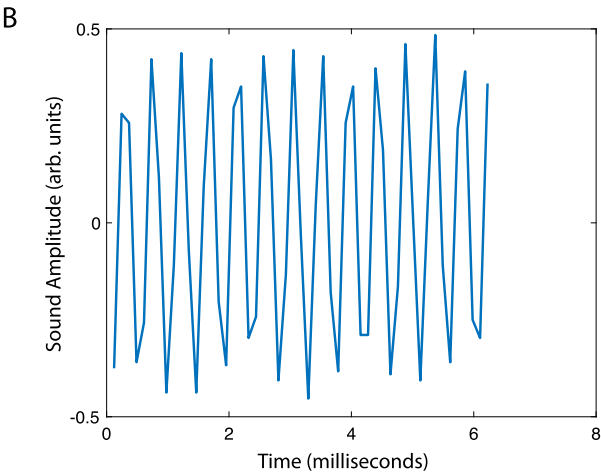
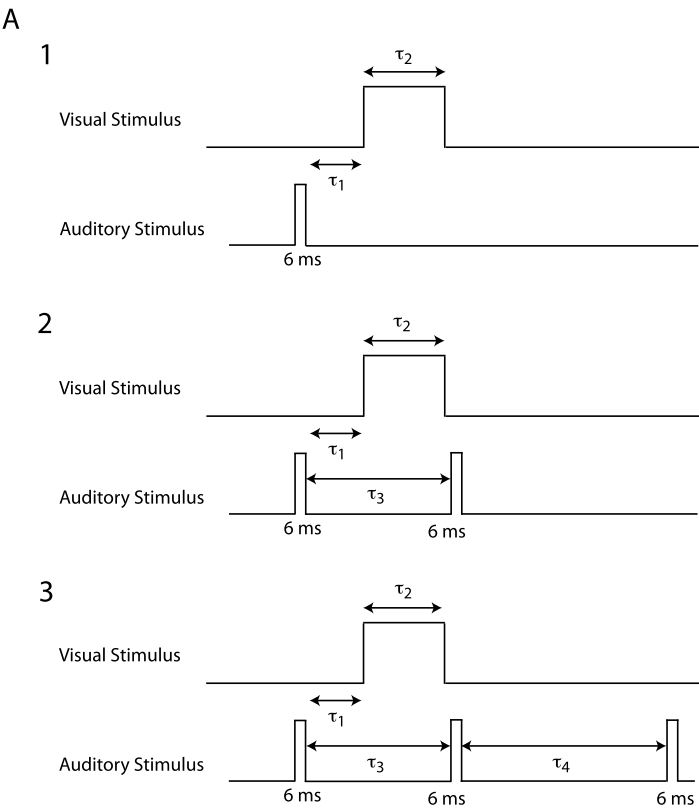
2.3.1. Experiment 1: Double Flash Illusory Dynamics

Participants were presented with a single visual flash of either a sharp-edged disk or a gradient disk, and with between 1 and 3 auditory beeps for each trial. The stimulus order was randomized among all of the possible visual and auditory stimulus combinations (6 combinations: 1 to 3 beeps for each of 2

potential visual stimuli (sharp-edged disk or gradient disk)). Participants performed 15 trials for each of the various stimulus combinations. Following the presentation of the crossmodal stimulus, participants were first asked to report the number of flashes perceived (1, 2, or 3 flashes; 3AFC). If the participant reported one flash perceived, they were also asked to choose the most appropriate description of the (real) flash that they perceived with reference to the three descriptions and plots shown in Fig. 4B (3AFC). Participants were instructed that the expansion then contraction alternative forced choice also included expansion only, and that the contraction then expansion alternative forced choice also included contraction only. If the participant reported two flashes perceived, they were asked to choose the most appropriate description of the two flashes (one real and one illusory) that they perceived with reference to the four descriptions and plots shown in Fig. 4C (4AFC). In this case, participants were instructed that each description and associated plot represented a set of related perceptions with one possible combination of dynamic and/or constant flashes (e.g., one dynamic flash followed by one constant flash). If the participant reported three flashes perceived, the number of perceived flashes was recorded, but no additional questions to elicit descriptions of the flashes were asked. The illusory dynamics of the flashed gradient disk stimuli were discovered during pilot observations by both the authors and naive participants, and the alternative forced choice options were generated based on these pilot observations. (Note: In order to avoid confusion, the alternative forced choice options (3AFC for one flash, and 4AFC for two flashes) were not randomized as presented in the sense that they were always displayed left to right in the same order in each case. All results for the gradient disk stimulus were compared with results for the sharp-edged disk control, which therefore should remove any bias generated by a lack of alternative forced choice randomization.)

2.3.2. *Experiment 2: Auditory Visual Binding With the Double Flash Illusion*

Participants were presented with a single simultaneous visual flash of either two sharp-edged disks, two gradient disks, or one of each (side by side), and with between 1 and 3 beeps for each trial. The stimulus order was randomized among all of the possible visual and auditory stimulus combinations (12 combinations: 1 to 3 beeps, with four potential visual stimuli (disk and disk, gradient and gradient, disk and gradient, or gradient and disk)). Participants performed 15 trials for each of the stimulus combinations. Following the presentation of the crossmodal stimulus, participants were first asked the number of flashes they perceived for the visual stimulus on the left (1 to 3 flashes, 3AFC), and then the number of flashes they perceived for the visual stimulus on the right (1 to 3 flashes, 3AFC). (Note: Full randomization of the stimulus locations prevented any sequence effect due to question order (i.e., inquiring



about the left stimulus and then the right stimulus). In addition, for the mixed stimulus combinations (gradient and disk, or disk and gradient) the number of flashes perceived for the gradient disk stimulus presented on the left was always combined with the number of flashes perceived for the gradient disk presented on the right. As a consequence, the question sequence of left then right was averaged out across stimulus locations.)

2.4. Statistics

The mean numbers of perceived flashes as well as the mean responses for the subsequent questions were calculated for each participant and then averaged across the group. We used two-tailed student *t*-tests (MATLAB function `ttest2`) for the statistical significance calculations reported below.

3. Results

3.1. Experiment 1: Double Flash Illusory Dynamics ($N = 10$)

Our first step was to confirm that the double flash illusion can occur with stimuli that are different than a sharp-edged disk, the stimulus used in the original double flash illusion. We determined the number of flashes perceived when either a sharp-edged disk or a gradient disk was presented in combination with 1 to 3 beeps, with all possible combinations randomly presented. For both visual stimuli, the numbers of flashes perceived with 1 flash and either 2 or 3 beeps were significantly greater than the number of flashes perceived with 1 flash and 1 beep ($N = 10$; $p < 0.01$ in all four cases; Fig. 4A). The number

Figure 3. Time Parameters of the Visual and Auditory Crossmodal Stimuli. The onset and offset times for visual and auditory stimuli are presented for Experiments 1 and 2 in this figure. Each panel represents a different stimulus condition: 1 flash and 1 beep (Panel A1), 1 flash and 2 beeps (Panel A2), and 1 flash and 3 beeps (Panel A3). The visual stimulus onset and offset times apply to the flashed visual stimulus (such as the sharp-edged disk or gradient disk), but not to the fixation cross, which is present for the entire trial duration. (Note: In Experiment 2, two visual stimuli are presented with the same presentation timing (onset–offset) and therefore are listed here as just one combined visual stimulus for simplicity.) The timing parameters are slightly different than those used in the original double flash illusion as a consequence of timing optimization to maximize the illusory effect with these particular visual stimuli. The timing parameters (τ_1 through τ_4) were measured with an oscilloscope five times for each stimulus type and then averaged. The time parameters for the sharp-edged disk and the gradient disk are presented in Table 1. Details on the measurement procedure are given in the Methods section. Panel B shows the auditory amplitude as a function of time for the beep stimulus. The auditory stimulus was played through both laptop speakers at equal amplitudes and at a sample rate of 8192 samples/second (0.122 ms per sample) within Psychtoolbox-3 using the MATLAB ‘sound’ function; the principal frequency component of each beep was 2.17 kHz. The background and auditory stimulus sound levels are given in the Methods section.

Table 1.
Sharp-edged and gradient disk stimulus mean timing parameters. Measured time durations of the auditory and visual stimuli, as well as times between stimuli presentations, are shown for both the sharp-edged and gradient disk cases. Time parameters were each measured five times with an oscilloscope and averaged (additional details are provided in the Methods section). The annotation 1B1F, for example, indicates that the stimulus contained one beep and one flash. The definition of each time parameter is indicated in Fig. 3.

Stimulus	Sharp-edged disk stimulus mean timing parameters (ms)			
	τ_1	τ_2	τ_3	τ_4
1B1F	17 ± 9	45 ± 9	N/A	N/A
2B1F	15 ± 9	34 ± 1	53 ± 7	N/A
3B1F	28 ± 5	41 ± 10	73 ± 13	104 ± 2

Stimulus	Gradient disk stimulus mean timing parameters (ms)			
	τ_1	τ_2	τ_3	τ_4
1B1F	39 ± 16	37 ± 7	N/A	N/A
2B1F	37 ± 13	34 ± 1	72 ± 12	N/A
3B1F	40 ± 9	33 ± 0	74 ± 9	104 ± 1

of flashes perceived with 1 flash and 3 beeps was larger than, but not significantly different than, the number of flashes perceived with 1 flash and 2 beeps ($N = 10$; $p > 0.32$ in both cases; Fig. 4A). This result is not surprising, given the extended delay between the second and third beeps, as discussed in more detail below. Furthermore, the number of flashes perceived for the gradient disk stimulus was not significantly different than the number of flashes perceived with the sharp-edged disk ($N = 10$; $p > 0.48$ in all three cases). This verifies that the gradient disk can generate the double flash illusion, and that it can do so with approximately the same relative frequency of occurrence as that of the original sharp-edged disk.

We next verified that the gradient disk initially induced dynamics when it was perceived to flash once (this includes the conditions in which 1 to 3 beeps accompanied the real flash, but no illusory flash was perceived). The dynamics that were perceived with the real flash are hypothesized to be the foundation for the dynamics perceived with the illusory flash, and therefore represent a critical step towards demonstrating dynamic replay. We found that the fraction of trials with dynamics reported (this includes the perception of expansion then contraction, expansion alone, contraction then expansion, or contraction alone, per our instructions to the participants) was significantly greater with the gradient disk than with the control stimulus (the sharp-edged disk) ($N = 8$; $t(14) = -4.22$, $p = 0.0009$; green and blue bars, Fig. 4B). These results

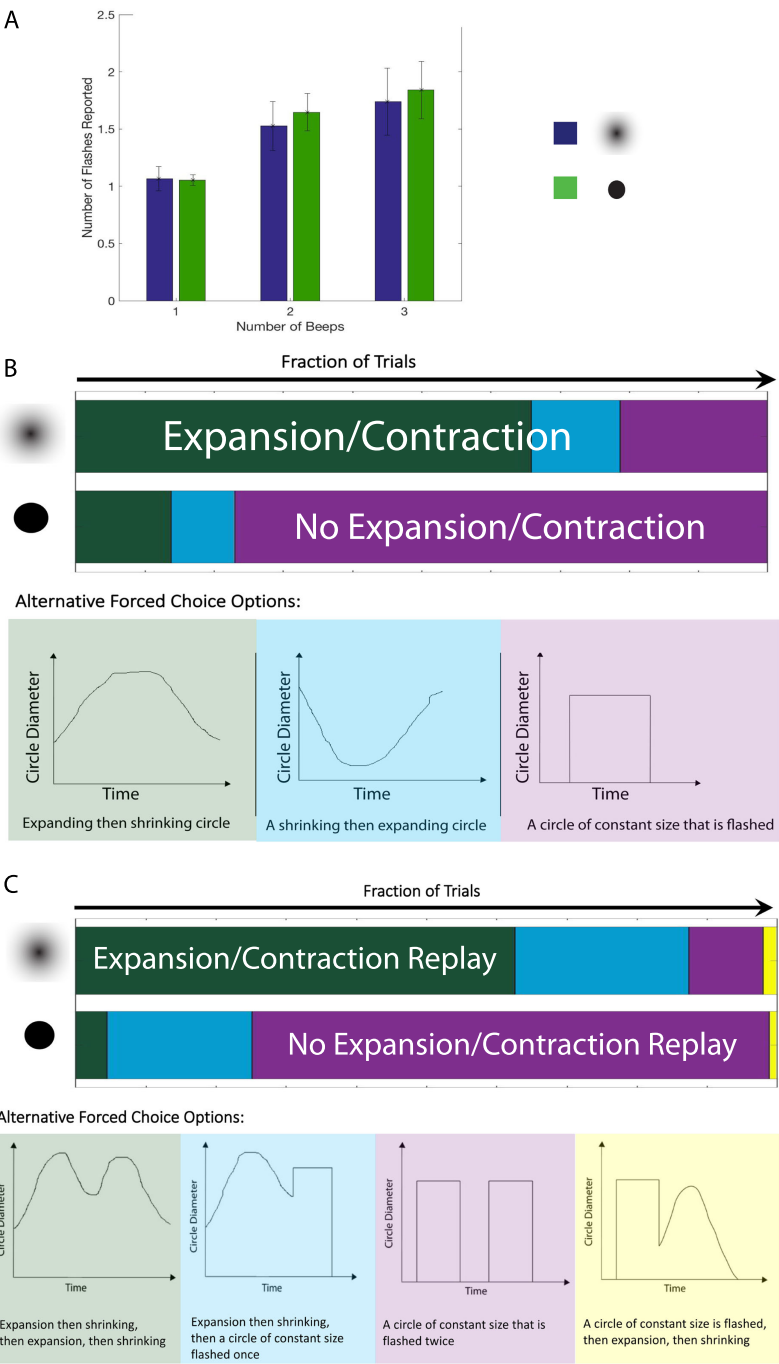
indicate that the gradient disk selectively generates the perception of illusory dynamics when one real flash is perceived.

The fraction of trials with dynamics reported for the gradient disk stimulus was compared to the fraction of trials with dynamics reported for the sharp-edged disk (rather than a baseline of chance) in order to verify that cognitive bias did not significantly affect the perception of dynamics recorded for the gradient disk stimulus. For example, a cognitive bias might arise from the presentation of the alternative forced choices in that a given participant might infer either that some form of perceived dynamics is likely, or that the ‘expansion then contraction’ alternative forced choice is the most likely as it is presented first in each case. This approach assumes that any cognitive bias would affect the sharp-edged disk and the gradient disk equivalently; therefore, the significant difference between the two stimuli helps eliminate any baseline bias that generates the reporting of illusory motion when none is actually perceived.

We finally tested whether the initial expansion (or contraction) associated with the real flash is also observed when an illusory flash is perceived. We selected the trials in which two flashes were perceived and asked participants if they perceived spatial dynamics in the first (real) flash, in the second (illusory) flash, in both the first and the second flash, or in neither. Participants were instructed to choose the first option in Fig. 4C (i.e., the alternative forced choice highlighted in green), if they perceived expansion and/or contraction (in any combination) for both the first and second flashes. The number of trials in which dynamics were perceived in both flashes was significantly greater for the gradient disk than for the control stimulus (sharp-edged disk) ($N = 8$; $t(14) = -2.47$, $p = 0.03$; green bar, Fig. 4C). In addition, the number of trials in which dynamics were perceived in both flashes (real and illusory) for the gradient disk was greater than the number of trials for the other three alternatives combined. As a consequence, if expansion (or contraction) dynamics were perceived with the first real flash for the gradient disk stimulus, then they were also most often perceived in the illusory flash as well.

3.2. Experiment 2: Auditory–Visual Binding With the Double Flash Illusion ($N = 11$)

We next investigated the differences observed in auditory–visual binding between the sharp-edged disk and the gradient disk. To test for binding, we used a modified version of the double flash illusion in which two flashed visual stimuli were presented simultaneously side by side with one, two, or three sequential beeps. The visual stimuli could be either two sharp-edged disks, two gradient disks, or one of each, randomized among all stimuli and also randomized left-right for the case of two dissimilar stimuli. Participants then responded with how many flashes they perceived for each of the two paired



stimuli in each trial, first the stimulus on the left, and then the stimulus on the right as discussed above.

When two *identical* visual stimuli were flashed with either two or three beeps (Fig. 5), the most common perceptions were either no beep-flash binding (i.e., only one flash was perceived for each visual stimulus), or binding to both visual stimuli (i.e., both visual stimuli were perceived to double or triple flash). In other words, the brain had a tendency to group the two identical visual stimuli together and either bind the sound(s) to both of the stimuli or bind to neither of them. In contrast, when two *different* visual stimuli (i.e., the gradient disk and the sharp-edged disk) were presented simultaneously with two or three beeps (as also illustrated in Fig. 5), the most common outcome was that the beeps bound to either one stimulus flash or the other but not to both (2 or 3 flashes of the gradient disk and 1 flash of the sharp-edged disk perceived side by side, or *vice versa*; i.e., differential binding). In other words, the beeps triggered one or more illusory flashes of one of the visual stimuli, but the other visual stimulus was not perceived to be followed by any illusory flashes.

When comparing the responses for the identical visual stimuli and the different visual stimuli conditions, the different visual stimuli had significantly greater differential binding (binding to one or the other stimulus but not both), than the identical visual stimuli ($N = 11$; $t(20) = 2.49$, $p = 0.02$). Therefore, the crossmodal network in the brain is selecting visual stimuli to bind with the auditory beeps differentially on a very short time scale. Furthermore, it is apparently selecting stimuli based on differences in their visual spatial properties (perhaps size, brightness, perceived dynamics, and spatial frequency spectrum, among others), which suggests a key parameter to consider in the perceptual crossmodal binding domain. [Note: This difference in crossmodal binding is not likely due to visuospatial attention. Participants were asked to fix their gaze on a fixation cross, and the visual stimuli were presented peripherally for a maximum of only 45 ms. There were no differences in the pre-stimulus

Figure 4. Dynamic Double Flash Perception (Experiment 1 Results). Panel A includes a plot of the number of flashes perceived when either a sharp-edged disk or a gradient disk is paired with one, two, or three beeps (the full length of each error bar is one standard deviation across participants). Participants also indicated the type of visual perception perceived when only a single flash was reported; this result is presented in Panel B for both the sharp-edged disk and the gradient disk (Stimuli: 1 real visual flash, and 1, 2, or 3 beeps; perceived: 1 flash). Note: The alternative forced choice color corresponds to the matching color in the bar plot (for example, the far left green alternative forced choice corresponds to the green fraction of the bar plot). Panel C indicates the temporal profile chosen when two flashes are perceived (Stimuli: 1 real visual flash, and 1, 2, or 3 beeps; perceived: 2 flashes). The alternative forced choice options shown in Panels B and C are the actual plots and descriptions shown to the participant during the experiment.

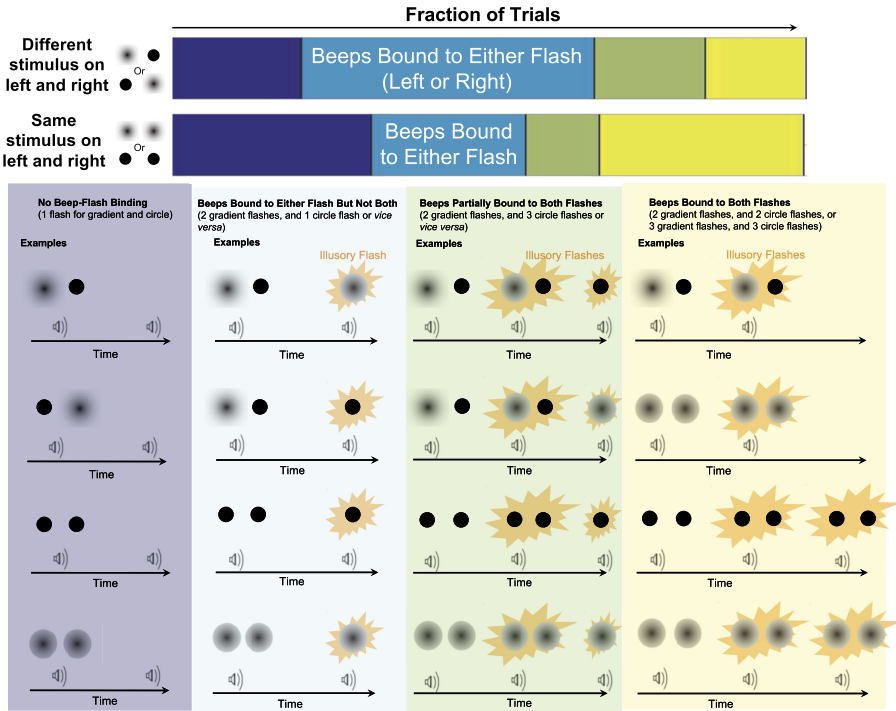


Figure 5. Differential Double Flash Binding of Two Simultaneous Visual Stimuli to Multiple Beeps (Experiment 2 Results). In this figure, a bar plot shows the visual perception reported when two visual stimuli are presented side by side with two or three beeps. The possible perceptions (alternative forced choices) are listed below the bar plot in colors that correspond to the subsection of the bar plot (and in the same order as the subsections of the bar plot). For example, the far left dark blue alternative forced choice corresponds to the far left dark blue fraction of the bar plot. The top bar plot indicates the perception reported when *different* stimuli are presented on the left and right (such as a sharp-edged disk next to a gradient disk), and the second bar plot indicates the responses when *identical* stimuli are presented on the left and right (such as two sharp-edged disks).

screen (display) between the identical visual stimuli cases and the different visual stimuli cases that would cause differences in pre-stimulus attention. Furthermore, the flashed visual stimuli were too short (45 ms) to allow for the orienting of attention (Koelewijn *et al.*, 2010).]

4. Discussion

4.1. Summary of Principal Results

In Experiment 1, we showed that a gradient disk visual stimulus generates illusory expansion and in some cases subsequent contraction when flashed, and that it was perceived to double and in some cases triple flash when presented

with either two beeps or three beeps. Furthermore, when the gradient disk was perceived to double flash (when accompanied by auditory beeps), dynamics were perceived in the second (illusory) flash that were similar to those observed in the first (real) flash. Thus the dynamic double flash illusion shows that the illusory flash is likely processed in ways that are similar to those of the real (stimulus) flash.

In Experiment 2, we presented a sharp-edged disk and a gradient disk simultaneously side by side to determine if they would bind differently to auditory beeps. We found that when the sharp-edged disk and the gradient disk were presented side by side, they were perceived to have more differential binding than when two identical stimuli were presented side by side. Therefore, we conclude not only that the double flash illusion can be used as a metric of auditory–visual binding, but also that visual-spatial properties (such as the presence of segregation, grouping, or binding within the visual modality) can impact crossmodal binding in the perceptual domain.

4.2. Possible Mechanisms for the Double Flash Illusion and Illusory Dynamics Replay

The double flash illusion generates reactivation of the visual cortex *via* auditory–visual connections, as triggered by an auditory stimulus. But why does the reactivation of early visual cortex generate a perceived repeat of the sharp-edged or gradient disk just seen, rather than of some other prior or random visual perception? One possible explanation may be that the residual pattern of the local network activation that is responsible for the first visual percept and its binding to the first sound is either reactivated, or the perceptual threshold of visibility itself is lowered, by the second sound. Based in part on the experiments described herein, one possible mechanism for the double flash illusion may be similar to the neural mechanism involved in Transcranial Magnetic Stimulation (TMS) replay (Jolij and Lamme, 2010; Liao *et al.*, 2013). In TMS replay studies, it has been proposed that as visual activation declines and passes from consciousness, subthreshold visual activation patterns are still present temporarily; these patterns can then be reactivated by a TMS pulse and brought back into visual awareness (i.e., TMS replay). Similarly, the double flash illusion could use auditory-triggered activation of visual cortex to reactivate latent patterns from recent visual stimuli, and then reprocess these reactivated patterns as new (illusory) perceptual experiences. This type of pattern reactivation in visual cortex is suggested by the dynamic double flash illusion and the illusory motion replay that participants report perceiving. Nonetheless, additional research is required to definitively prove this hypothesis. Future experiments that may be able to further support this hypothesis are described below.

One key question that might be asked is why participants occasionally see a single flash (no illusory flash), even with the double flash auditory–visual stimulus (i.e., one flash and two beeps). First and foremost, this observation is consistent with those of the original double flash illusion (Shams *et al.*, 2000, 2002). Variability in illusion occurrence could be due to variability in the strength of the visual cortical activation from the first real flash (modulated by effects such as attention). In addition, crossmodal connectivity and phase synchrony before the illusion, determined by EEG and MEG measures, predict the probability of the double flash illusion occurring (Keil *et al.*, 2013; Lange *et al.*, 2011), and are likely major factors in the perception of the illusory flash. In other words, whether the visual stimulus can be reactivated depends on how strong the initial visually-driven activation is (i.e., how strong visual activation remains after the percept of the real flash has left consciousness, and as a consequence how much of a boost is needed for the illusory flash to reach consciousness). In addition, the occurrence of the double flash illusion depends on how strong the multimodal activation of visual regions is from auditory brain regions. If one or both of these factors is too weak, the second illusory flash will likely not occur. As these factors vary from individual to individual, and also from moment to moment within a given individual, variability in the number of double flashes perceived within and across individuals might be expected.

A corollary question is why subjects sometimes see a first real flash with illusory motion, followed by a second illusory flash without the replay of illusory motion. We hypothesize that when the illusory motion is not replicated in the illusory flash, the re-activation of the latent pattern is too weak to fully represent the spatial pattern that triggers the illusory motion. Therefore, only the highest amplitude part of the pattern becomes supra-threshold (i.e., the gradient disk center) and no illusory motion is perceived. It follows that one could see a fragment of the flashed pattern (i.e., a smaller central gradient disk) that has weakened dynamics or has no dynamics at all. This could in theory be tested by asking participants the approximate size of the second flash when the first flash is perceived to have dynamics and the second flash does not.

An additional question of interest is why there are more, but not significantly more, flashes perceived with three beeps rather than two beeps. With reference to Fig. 4A, for the sharp-edged disk 1.65 flashes were perceived for the 1 flash 2 beep condition, and 1.84 flashes were perceived for the 1 flash 3 beep condition; for the gradient disk, 1.53 flashes were perceived for the 1 flash 2 beep condition, and 1.74 flashes were perceived for the 1 flash 3 beep condition. The three beep condition likely creates only a few additional replays beyond one replay because the first replay (illusory flash) is typically weaker than the original stimulus activation. In other words, illusory flashes are not fully regenerative of previous flashes, as they result from a smaller yet

still supra-threshold activation. In order to perceive a second illusory flash (or third total flash) *via* reactivation of visual cortex, an even larger boost in activation from auditory regions would be required to surpass threshold, thereby generating a second replay. In this experiment, the delay between the second beep and the third beep was approximately 104 ms, considerably longer than the delay between the first beep and the second beep. This extra delay could also reduce the number of triple flashes perceived by participants.

From this perspective, the perception of illusory motion, just like the perception of illusory flashes, likely has diminishing probability with an increasing number of beeps due to iterative loss in visual cortical activation level with each reactivation. In particular, the decrease in visual activation associated with the illusory flash means that this flash will likely be lower in amplitude than the real flash, and will therefore be characterized by a smaller gradient with consequently less illusory motion. This hypothesis may be testable with an EEG experiment that is designed to measure the relationship between the temporal evolution of visual cortical activation and the number of replays perceived.

Additional research related to the replay hypothesis is also of interest to further examine the differential effects of gradient shape on the illusory motion perceived. We have already piloted the visual illusory motion perception of a variety of gradients, and have found associated variations of the visual illusory motion space-time dynamics depending on the shape of the brightness gradient over space. We also have informally observed a trade-off between larger sized stimuli, which exhibit more perceived dynamics, and smaller sized stimuli, which exhibit more illusory flashes. We hypothesize that the stronger the illusory motion perceived in the initial stimuli is, the more likely it is that there will be a subsequent replay of that illusory motion. However, the relationship between the presence of the illusory flash and the perception of the illusory dynamics should be further examined in follow-up work. For example, additional tests with EEG or fMRI may be able to relate the strength of visual activation in primary regions and visual motion regions to the perception of dynamics associated with two or more beeps.

4.3. Possible Mechanisms for the Generation of Illusory Dynamics

The experiments described herein suggest that not only the illusory flash, but also its dynamics, are processed in a way that is similar to that of a real flash with *its* associated (illusory) dynamics. We consider herein two possible mechanisms for the illusory dynamics associated with the real flash to occur: (1) the eventual perception of the visual dynamical behavior is initiated early in the visual processing pipeline by an expanding wave of activation beginning at the gradient center, or (2) a static visual pattern is generated by the gradient flash in early visual cortex, but the illusory dynamics are not generated until higher

visual cortex. It appears most likely that the second beep simply re-triggers the mechanism that generates the illusory visual dynamics associated with the real flash.

The gradient disk flash likely generates illusory visual dynamics through differences in how rapidly high contrasts and low contrasts are processed (Backus and Oruç, 2005, and references cited therein). The effect of contrast on the perception of motion was studied extensively by Seiffert and Cavanagh, who argued that this effect is due to a combination of low level motion-energy and velocity detection mechanisms (Seiffert and Cavanagh, 1999). This points toward the dynamic double flash being processed by the first possible mechanism described above, in this case by a low-level illusory motion generator. If the first (bottom up) mechanism generates the dynamics of the illusory motion, this supports the hypothesis that a feedforward, low-level process generates the double flash illusion (which is consistent with previous EEG and fMRI studies (Shams *et al.*, 2001, 2005)).

From this perspective, the sudden presentation of the dark gradient disk against a uniform white background initially generates maximum contrast at the center of the gradient, with decreasing contrast toward the edges. Hence, if the center is perceived first, followed progressively by the remainder of the pattern, then an expanding disk will be perceived. Some participants occasionally commented that they perceived first expansion and then contraction for the gradient disk stimulus, the latter perception likely arising from stimulus offset. This case is more difficult to analyze, as the background for stimulus offset is not uniform, but is instead the gradient disk just presented. This patterned background then transitions suddenly to a uniform background at stimulus offset. Given the relatively short time between onset and offset, and given further that the stimulus is presented peripherally, complicated offset dynamics (such as the perception of contraction following the initial expansion) are not surprising.

4.4. *Possible Mechanisms for the Generation of Differential Binding*

In this study we also explored the illusory perception generated from the presentation of two visual stimuli side by side, accompanied by two beeps. This experiment tested for differential binding due to visual object grouping with multiple simultaneous visual stimuli. It is a somewhat surprising finding from the perspective of neural processing speed that the visual cortex can differentially select visual stimuli for crossmodal binding in the short time period of the double flash illusion. This result implies both rapid visual segregation of the stimuli and selection between the visual stimuli prior to their binding to the auditory stimuli. The binding principle may be simple, in that the visual stimulus with the strongest early visual activation is bound to the auditory

stimulus, or it may be more complex, engendered by a visuospatial metric that depends on the individual's auditory–visual network.

4.5. *Implications for Crossmodal Binding*

Previous investigations of crossmodal illusions have already determined several additional temporal and spatial properties that mediate auditory–visual stimulus binding. In particular, auditory–visual simultaneity was found to be an important feature for binding in the bouncing *vs.* streaming illusion (in which two approaching visual objects are perceived to bounce, rather than stream, when a short beep is timed to play at the moment of their intersection) (Sekuler, 1997). Watanabe and Shimojo (2001) further elucidated the role of simultaneity in the bouncing *vs.* streaming illusion by determining that the beep onset timing was required to be within 300 ms of the visual intersection for bouncing (or binding) to be perceived.

Although stronger auditory–visual spatial congruence was not observed to enhance the frequency of double flash illusion perception (Innes-Brown and Crewther, 2009; Kumpik *et al.*, 2014), auditory–visual binding behavior can be affected by similarity in stimulus spatial location, as highlighted by the ventriloquist effect (in which the perceived location of an auditory beep shifts toward a temporally-associated visual stimulus). In this illusion, larger percentage shifts (the percentage of the auditory–visual disparity that the auditory beep shifts toward the visual flash) or stronger binding occur when the auditory beep is closer in space to the visual stimulus (Hairston *et al.*, 2003). Therefore, simultaneity and co-localization are two known stimulus properties that can affect the crossmodal binding of stimuli.

In our differential binding experiment (Experiment 2), we took these auditory–visual binding principles of simultaneity and co-localization into account by determining whether simultaneously presented and co-localized but non-identical visual stimuli can be bound differentially to auditory stimuli *via* the double flash illusion paradigm (with the auditory stimuli placed equidistant from both visual stimuli). In particular, our results indicate the significant role of visuospatial patterns, edges, and gradients in the selection and segregation of visual stimuli for crossmodal binding.

4.6. *The Dynamic Double Flash Illusion as a Metric of Neural Function and Dysfunction*

The double flash illusion has served as an effective metric for auditory–visual integration dysfunction in diseases such as schizophrenia and autism, as well as in normal development and aging. Patients suffering from schizophrenia and autism have been shown to have a larger tolerance for auditory–visual asynchrony (window of integration) in the double flash illusion (Haß *et al.*, 2017) (Zhou *et al.*, 2018) than neurotypical individuals, while synaesthetes

have been found to have a smaller tolerance (Neufeld *et al.*, 2012). In addition, older individuals (65+ years) have been shown to have a wider integration window than younger participants (18 to 30 years old) (McGovern *et al.*, 2014). Such differences in auditory–visual integration highlight broad cross-modal sensory changes in these conditions that are just beginning to be studied and understood.

The dynamic double flash illusion as well as other double flash modifications, such as the illusory audiovisual rabbit and invisible audiovisual rabbit illusions, may provide useful complementarity to the standard double flash illusion as a metric of auditory–visual processing (Stiles *et al.*, 2018). For example, the dynamic and spatial elements of the dynamic double flash illusion could provide additional insight into the crossmodal integration that characterizes various disease states. If a dysfunctional sensory system elicits a weaker dynamic double flash illusion than the original double flash illusion, this may imply a difficulty in coupling more complex visuospatial patterns to auditory perception. In addition, the audiovisual rabbit illusions exhibit the unusual property of postdiction, in which subsequent events influence the perception of prior events. An inability to process postdictive stimuli could be an indication of a difficulty in providing contemporaneous feedback in conscious perception.

5. Conclusion

The dynamic double flash illusion demonstrates that a sequence of auditory tones co-presented with a single real visual flash can generate not only an illusory flash, but also an illusory flash with spatial characteristics that are similar to those of the real flash. In addition, the generation of illusory dynamics by spatial gradients in the real flash can generate the perception of similar illusory dynamics in the illusory flash. Simultaneous presentation of real visual flashes with different spatial characteristics can exhibit differential binding to a pair of sequential auditory tones. Further study of the dynamic double flash illusion may allow differentiation among various models of crossmodal sensory integration and processing. Finally, the application of this new dynamic double flash illusion to the study of disease states could be helpful in further elucidating the effects of dysfunction and disease on crossmodal sensory perception and integration.

Acknowledgements

We are pleased to acknowledge key insights and useful comments from Daw-An Wu, the associate editor, and the reviewers. We are also grateful for support

from the National Institutes of Health; the Philanthropic Educational Organization Scholar Award Program; the Arnold O. Beckman Postdoctoral Scholars Fellowship Program; the National Science Foundation; and the Core Research for Evolutional Science and Technology Program of the Japan Science and Technology Agency.

Conflicts of Interest

The authors declare no competing financial interests.

Supplementary Material

Supplementary material is available online at:
<https://doi.org/10.6084/m9.figshare.9164189>

References

- Backus, B. T. and Oruç, I. (2005). Illusory motion from change over time in the response to contrast and luminance, *Journal of Vision* **5**, 1055–1069.
- Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L. and Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias, *Journal of Cognitive Neuroscience* **15**, 20–29.
- Haß, K., Sinke, C., Reese, T., Roy, M., Wiswede, D., Dillo, W., Oranje, B. and Szycik, G. R. (2017). Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion, *Cognitive Neuropsychiatry* **22**, 145–158.
- Innes-Brown, H. and Crewther, D. (2009). The impact of spatial incongruence on an auditory–visual illusion, *PLOS One* **4**, e6450.
- Jolij, J. and Lamme, V. A. (2010). Transcranial magnetic stimulation-induced ‘visual echoes’ are generated in early visual cortex, *Neuroscience Letters* **484**, 178–181.
- Keil, J., Müller, N., Hartmann, T. and Weisz, N. (2013). Prestimulus beta power and phase synchrony influence the sound-induced flash illusion, *Cerebral Cortex* **24**, 1278–1288.
- Koelewijn, T., Bronkhorst, A. and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies, *Acta Psychologica* **134**, 372–384.
- Kumpik, D. P., Roberts, H. E., King, A. J. and Bizley, J. K. (2014). Visual sensitivity is a stronger determinant of illusory processes than auditory cue parameters in the sound-induced flash illusion, *Journal of Vision* **14**, 12–12.
- Lange, J., Oostenveld, R. and Fries, P. (2011). Perception of the touch-induced visual double-flash illusion correlates with changes of rhythmic neuronal activity in human visual and somatosensory areas, *Neuroimage* **54**, 1395–1405.
- Liao, H.-I., Wu, D.-A., Halelamien, N. and Shimojo, S. (2013). Cortical stimulation consolidates and reactivates visual experience: neural plasticity from magnetic entrainment of visual activity, *Scientific Reports* **3**, 2228.
- McGovern, D. P., Roudaia, E., Stapleton, J., McGinnity, T. M. and Newell, F. N. (2014). The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration, *Frontiers in Aging Neuroscience* **6**, 250.

- Mishra, J., Martinez, A., Sejnowski, T. J. and Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion, *Journal of Neuroscience* **27**, 4120–4131.
- Neufeld, J., Sinke, C., Zedler, M., Emrich, H. M. and Szyzik, G. R. (2012). Reduced audio-visual integration in synaesthetes indicated by the double-flash illusion, *Brain Research* **1473**, 78–86.
- Seiffert, A. E. and Cavanagh, P. (1999). Position-based motion perception for color and texture stimuli: Effects of contrast and speed, *Vision Research* **39**, 4172–4185.
- Sekuler, R. (1997). Sound alters visual motion perception, *Nature* **385**, 308–308.
- Shams, L., Iwaki, S., Chawla, A. and Bhattacharya, J. (2005). Early modulation of visual cortex by sound: An MEG study, *Neuroscience Letters* **378**, 76–81.
- Shams, L., Kamitani, Y. and Shimojo, S. (2000). What you see is what you hear, *Nature* **408**, 788.
- Shams, L., Kamitani, Y. and Shimojo, S. (2002). Visual illusion induced by sound, *Cognitive Brain Research* **14**, 147–152.
- Shams, L., Kamitani, Y., Thompson, S. and Shimojo, S. (2001). Sound alters visual evoked potentials in humans, *NeuroReport* **12**, 3849–3852.
- Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions, *Current Opinion in Neurobiology* **11**, 505–509.
- Stiles, N. R. B., Li, M., Levitan, C. A., Kamitani, Y. and Shimojo, S. (2018). What you see is what you will hear: Two new illusions with audiovisual postdictive effects, *PLOS One* **13**, e0204217.
- Watanabe, K. and Shimojo, S. (2001). When sound affects vision: effects of auditory grouping on visual motion perception, *Psychological Science* **12**, 109–116.
- Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D. and Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception, *Neuroimage* **31**, 1247–1256.
- Zhou, H.-Y., Cai, X.-L., Weigl, M., Bang, P., Cheung, E. F. and Chan, R. C. (2018). Multi-sensory temporal binding window in autism spectrum disorders and schizophrenia spectrum disorders: A systematic review and meta-analysis, *Neuroscience & Biobehavioral Reviews* **86**, 66–76.